ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

Optimal Control Model of Technology Penetration

Donald A. Hanson, Yaroslav Kryukov, Sven Leyffer, and Todd S. Munson

Mathematics and Computer Science Division

Preprint ANL/MCS-P1544-0908

September 22, 2008

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

Contents

Optimal Control Model of Technology Transition*

Donald A. Hanson,† Yaroslav Kryukov,‡ Sven Leyffer, and Todd S. Munson§

September 25, 2008

Abstract

This paper discusses the use of optimization software to solve an optimal control problem arising in the modeling of technology penetration. We set up a series of increasingly complex models with such features as learning by doing, adjustment cost, and capital investment. The models are written in continuous time and then discretized by using different methods to transform them into large-scale nonlinear programs. We use a modeling language and numerical optimization methods to solve the optimization problem. Our results are consistent with findings in the literature and highlight the impact the discretization choice has on the solution and accuracy.

1 Introduction

Our goal is to compute an optimal transition from conventional (old) to low-emission (new) technology for energy production. The new technology has higher costs but a lower emission rate of greenhouse gases, making it possible to reduce emissions without substantial reductions in energy consumption that would be necessary using only the old technology.

We are interested in the socially optimal output schedules of both technologies; these tell us the best possible scenario that could be achieved if the entire energy industry were controlled by a (benevolent and omniscient) single agency. In reality, the energy industry consists of many independent firms, which have to be motivated by policy measures to adopt the new technology. Our model could generate several important inputs to construct such a policy. First, our model determines the optimal output schedule, or transition path, that serves as the ultimate goal of the policy. Second, our model can be used to compute the exact amount of policy intervention (tax rate, emission quota, etc.) Third, if a policy fails to achieve the optimal transition, our model can show how large the shortfall is and whether correcting it justifies the additional effort.

We seek to develop a model that provides a realistic transition path; in other words, the output of either technology should be a continuous, but not necessarily smooth, function of time. In addition, the total energy output should be increasing over time.

*Preprint ANL/MCS-P1544-0908

†Decision and Information Sciences, Argonne National Laboratory dhanson@anl.gov

‡Tepper School of Business, Carnegie Mellon University, kryukov@andrew.cmu.edu

§Mathematics and Computer Science Division, Argonne National Laboratory, {leyffer,tmunson}@mcs.anl.gov

The increasing energy output is considered necessary given the continued growth of world population and economic well-being. Gradual transition is motivated by historic data on actual technology transitions that find the penetration rate of the new technology to be an S-shaped function of time—although it is increasing throughout the entire transition period, the penetration rate is convex during the early stages of transition, then passes through an inflection point, and turns concave as it approaches full adoption; see Jensen (1982) and Geroski (2000). To be consistent with these results, we develop models with various features such as learning by doing in the new technology, which reduces the unit cost as the cumulative output increases; transition costs, which penalize fast changes in output; and capital investment.

Our paper considers the following economic concepts: technology diffusion, environmental policy, and learning by doing. Popp (2004) provides a concise overview of these concepts. The diffusion (adoption) literature studies the technology adoption by individual firms and focuses on factors that cause firms to adopt at different times, and the shapes of transition paths observed in the past. Jensen (1982) quotes a number of studies that find the initial (convex) stage to be relatively short; sometimes it is not present at all, leading to a kink in the adoption path when the adoption starts and a concave shape after that. Cabral (1990) concentrates on the inflection point of an S-shaped function, arguing that it is an approximation to a discontinuous jump.

Jensen (1982, 1983) assumes some firms need more time than others to convince themselves that adoption is worthwhile; their model generates both S-shaped and concave transition paths. Tonks (1986) has a consumer cautiously discovering a new good, leading to a concave adoption path that overshoots the long-term asymptote. Cabral (1990) assumes some firms will not join until they see adoption by a certain share of the market; he generates discontinuities (vertical segments) in the adoption path. Balcer and Lippman (1984) assume multiple generations of a technology, whereas a firm will not adopt if a new technology does not offer a substantial improvement over the already installed one or if a better technology is expected to be released soon. Abstracting from economics, Geroski (2000) discusses diffusion models based on ideas from epidemiology, sociology, and information dissemination. Benhabib and Spiegel (2005) look at the diffusion of general technological knowledge between countries rather than firms; they find different rates of convergence to the "world leader" and explain them by different levels of education.

A related area of economic research studies environmental policy, specifically government measures aimed at inducing the firms to behave in a socially optimal way. Milliman and Price (1989) demonstrate (in theory) how frequently suggested policies such as emission taxes, subsidies, and tradable permits induce innovation and diffusion of cleaner technology. Newell et al. (2006) provide empirical evidence supporting both the effectiveness of such policy measures and profit maximizing behavior predicted by the diffusion literature.

Woerlen (2004) provides an overview of learning by doing, discussing both the theoretical issues and evidence of the experience-driven reduction in capital cost of the alternative energy.

Our results contribute to this literature on both the economic and the methodological side. The different models we develop demonstrate different forms of transition behavior, suggesting that our optimal transition paths are potentially implementable by policy intervention. From an economic standpoint, we demonstrate that the realistic gradual transition can be generated without resorting to multiple agents and is, in fact, socially optimal under reasonable assumptions. We find that learning by doing alone leads to discontinuous instant transition; adding adjustment costs ensures continuous and smooth transition. Our capital investment model results in a transition path that is continuous and concave after the initial kink.

From a methodological standpoint, our paper offers several important insights. First, we re-examine the use of discrete-time models to approximate the continuous-time model. A typical economic paper assumes discrete time with a period of one year. This decision is driven by the annual nature of input data and an assumption that the length of the period does not affect results. We adopt a more cautious approach: we formulate the model in continuous time, then discretize it using several discretization methods, and vary the period length until all discretization methods converge. In at least one case, we demonstrate how a large period length combined with a discretization method commonly used in economics can lead to a solution that differs from the true solution in a fundamental way. Second, we use modern computational tools to solve our dynamic problems. In the studies discussed above, the optimization is performed analytically, which is possible only for highly stylized models that do not describe actual industries. Numerical methods overcome this constraint and allow us to solve more realistic problems.

In the past, the limited power of early computers and the large scale of dynamic problems forced computational economists to use indirect approaches, most notably replacing an optimization problem with a system of optimality conditions, and applying Gauss-Jacobi or Gauss-Seidel iterations of updating the vector of variables one component at a time. However, these methods are notoriously slow and unreliable. Moreover, using (first-order) optimality conditions requires careful assumptions about curvature. We solve the original optimization problem instead, avoiding the need both to derive optimality conditions and to implement the solution algorithm. The combination of modern computers and optimization algorithms allows us to solve problems with thousands of variables in seconds.

Moreover, we demonstrate the ease with which a high-level modeling language such as AMPL (Fourer et al., 2003), can be used to experiment with the models. Unlike programming languages (C, Fortran), the modeling language allows the users to specify the model in its original algebraic form.

The rest of the paper is structured as follows. Section 2 develops the various models, Section 3 explains the discretization method, Section 4 presents the results, and Section 5 discusses possible extensions.

2 Model Description and Background

We solve the social planner's problem of maximizing the social welfare, with the condition that total accumulated emissions at a certain point in time will not exceed the specified level. Our variables are energy output schedules of old and new technology, where the new technology has lower emissions but is more expensive, although its costs are expected to decrease with wider adoption. We formulate a number of models with increasing detail and features.

2.1 Common Economic Components and Parameters

We first define the components common to all three models. Specific functional forms are chosen for demonstration; however, our approach is not limited to these, and other choices are possible.

Time. All our model are dynamic, with continuous time and finite horizon: $t \in [0, T]$. We denote functions of time as $x(t)$ and their derivatives as

$$\dot{x}(t) = \frac{dx(t)}{dt}.$$

We use continuous discounting with the rate $r > 0$.

Energy Output. There are two technologies, old and new, and their energy output at time t is denoted $q^o(t)$ and $q^n(t)$, respectively. We also define the total output to be $Q(t) = q^o(t) + q^n(t)$.

Demand and Consumer Surplus. The benefit of energy to society is represented by the consumer surplus $\tilde{S}(Q, t)$, computed as the integral of demand and scaled by the demand growth rate (hence the dependence on time). In our models, we use the following functional form for the consumer surplus:

$$\tilde{S}(Q, t) = e^{bt} S(Qe^{-bt}), \tag{2.1}$$

where $b > 0$ is the growth rate of demand. The consumer surplus is derived from the constant elasticity of substitution (CES) utility, as described in Appendix A:

$$S(Q) = \int p(q)dq\bigg|_{q=Q} = \int \frac{S_0}{q^\sigma}dq\bigg|_{q=Q} = \begin{cases} S(Q) = S_0 \ln Q, & \text{if } \sigma = 1 \\ \frac{S_0}{1-\sigma} Q^{1-\sigma}, & \text{otherwise,} \end{cases}$$

where $\sigma > 0$ is the demand parameter. The functional form of (2.1) is due to the fact that the growth factor e^{bt} is applied to the direct demand function $q(p)$ rather than the inverse demand $p(q)$.

Production Costs. We assume constant marginal costs. Each unit of energy is produced with the old technology costs c_o. The cost of the new technology is subject to learning by doing; that is, the unit cost $c_n(x(t))$ is a decreasing function of cumulative output, which we define as

$$x(t) = \int_0^t q^n(\tau)d\tau. \tag{2.2}$$

Following the economic literature (Woerlen, 2004) and discussion in Appendix A, we let

$$c_n(x) = c_n^0 \left[\frac{x}{\bar{X}} + 1\right]^{\log_2 \gamma}, \tag{2.3}$$

where the parameters $\bar{X}$ and γ are described in Table 1.

Greenhouse Gases Emissions. Producing energy generates greenhouse gases at the unit rate of $b_o > 0$ with the old technology and $b_n \in (0, b_o)$ with the new technology. We are interested in limiting cumulative emissions at the end of the modeling period. Since earlier emissions do more damage (for example, by the irreversible reduction of glaciers), we discount the emissions at the environmental time preference rate a. We use $a \in (0, r)$, but there also exist economic justifications for $a = 0$ or $a > r$. The constraint on cumulative emission is

$$\int_0^T e^{-at}\big(b_o q^o(t) + b_n q^n(t)\big)dt \le z_T. \tag{2.4}$$

Units and Parameters. We make our definitions more precise by setting specific units and parameter values. The quantities ($q(t)$'s, $x(t)$) are in "quads," or billion of millions (10^{15}) of BTUs[1]; monetary amounts (objective, $S(Q)$, etc.) are in billions of dollars; and emissions are measured in billion tons of carbon (tC). See Table 1.

Table 1: Parameter values common to all models

Parameter	Unit	Notation	Value
Discount rate	-	r	0.05
Demand exponent	-	σ	2.0
Demand scale	\$B	S_0	98,000
Demand growth rate	-	b	0.015
Environmental rate	-	a	0.02
Emissions, old tech.	tC/mBTU	b_o	0.02
Emissions, new tech.	tC/mBTU	b_n	0.001
Unconstrained emissions	BtC	$Z_{\max}$	61.9358
Emission reduction %	-	ζ	0.5
Production cost, old tech.	\$/mBTU	c_o	20
Starting cost, new tech.	\$/mBTU	c_n^0	50
Learning rate	-	γ	0.85
Initial experience	quad	x_0	0
Experience unit size	quad	$\bar{X}$	300

The emission cap is computed as $z_T = (1-\zeta)Z_{max}$, where Z_{max} equals the cumulative emissions when they are not constrained (which naturally leads to zero utilization of the new technology). The cost parameters γ and $\bar{X}$ were selected to achieve $c_n(x(T)) \approx 30$, a 60% reduction in unit cost

[1]British thermal unit (BTU) is the unit of energy used in power industry. It is equal to 1,055 joules.

by the end of the modeling period (but still more expensive than the old technology with its unit cost of 20). The demand scale is calibrated to match current output level: $q^o(0) + q^n(0) = 60$.

2.2 Model I: Basic Model

The first model takes into account only the effect of learning by doing and selects the energy output schedules to maximize the net discounted welfare without exceeding the emission cap. The model is an optimal control problem. Energy output amounts $q^o(t), q^n(t)$ are the controls, and the state variables are the experience level $x(t)$ and the accumulated emissions $z(t)$:

$$\begin{aligned}
&\underset{\{q^o,q^n,x,z\}(t)}{\text{maximize}} && \int_0^T e^{-rt}\left[\tilde{S}(q^o(t)+q^n(t),t) - c_o q^o(t) - c_n(x(t))q^n(t)\right] dt && \text{(2.5a)}\\
&\text{subject to} && \dot{x}(t) = q^n(t), \qquad x(0) = x_0 = 0 && \text{(2.5b)}\\
& && \dot{z}(t) = e^{-at}\big(b_o q^o(t) + b_n q^n(t)\big), \qquad z(0) = z_0 = 0 && \text{(2.5c)}\\
& && z(T) \le z_T && \text{(2.5d)}\\
& && q^o(t) \ge 0, \qquad q^n(t) \ge 0. && \text{(2.5e)}
\end{aligned}$$

The objective (2.5a) is the discounted net welfare, computed as the difference between consumer surplus and production cost. Constraint (2.5b) defines the cumulative output $x(t)$ and is a transformation of (2.2) into a differential equation; this transformation has the advantage that the discretized equations are sparse, allowing us to use the large-scale nonlinear programming (NLP) solvers. Constraints (2.5c) and (2.5d) represent the cumulative emission cap (2.4); (2.5c) is again a differential constraint that defines $z(t)$ to be cumulative emissions at time t, and (2.5d) imposes the cap. Constraint (2.5e) requires that output amounts be nonnegative. We do not need to impose a nonnegativity constraint on $x(t)$ because (2.5b) and (2.5e) guarantee it.

2.3 Model II: Adjustment Cost

Model I is rather simplistic and does not take adjustment costs into account. As a consequence, the resulting transition shows an instantaneous switch, or bang-bang control. Hence, we refine the model by imposing an adjustment cost on the increase in new energy output, $q^n(t)$. Keeping the rest of model components the same, we define

$$y(t) = \max\{0, \dot{q}^n(t)\}, \tag{2.6}$$

the positive change in $q^n(t)$. We use it to penalize the rate of change in the output of the new technology by subtracting the adjustment cost $c_y y^2(t)$ from the objective. The power on $y(t)$ makes the adjustment cost convex to reflect the difficulty of rapid change. The only new parameter is the scale of the adjustment cost, which we set as $c_y = 10$. The model is given by

$$\underset{\{q^o,q^n,x,y\}(t)}{\text{maximize}} \quad \int_0^T e^{-rt}\left[\tilde{S}(q^o(t)+q^n(t),t) - c_o q^o(t) - c_n(x(t))q^n(t) - c_y y^2(t)\right] dt \tag{2.7a}$$

$$\text{subject to} \quad \dot{x}(t) = q^n(t), \qquad x(0) = x_0 = 0 \tag{2.7b}$$

$$\dot{q}^n(t) \le y(t), \quad q^n(0) = 0 \tag{2.7c}$$

$$0 \le y(t) \tag{2.7d}$$

$$\dot{z}(t) = e^{-at}[b_o q^o(t) + b_n q^n(t)], \qquad z(0) = z_0 = 0 \tag{2.7e}$$

$$z(T) \le z_T \tag{2.7f}$$

$$q^o(t) \ge 0, \qquad q^n(t) \ge 0. \tag{2.7g}$$

We retain all the model components from the previous subsection; objective (2.7a) now includes the adjustment cost term $c_y y^2(t)$.

The new constraints are (2.7c) and (2.7d), which are an equivalent way of stating (2.6) because the objective term ensures that at least one of the constraints (2.7c)–(2.7d) does bind. The initial condition on $q^n(0)$ is necessary for a well-defined problem but does not affect the solution because we have $q^n(t) = 0$ for sufficiently small t's.

2.4 Model III: Capital Investment

The penalty on the rate of increases in output is a somewhat artificial construct. The resulting adjustment cost was motivated by the need to build up capacity in the new technology. Here, we model this capital investment directly. To simplify the notation, we index the technologies by $j \in \{o, n\}$.

Capital and Investment. $K^j(t)$ is the amount of capital in technology j at time t. It is increased by an investment $I^j(t)$ and depreciates at a constant annual rate of $\delta \in (0,1)$. We denote the initial capital level as $\bar{K}_0^j$.

Costs. Instead of the constant unit costs (cq) used in the previous models, $C^j(q,K)$ now denotes the total cost of producing energy quantity q with technology j and capital K. We expect $C^j(q,K)$ to be increasing and possibly convex in q (increasing marginal costs) and decreasing and concave in K (decreasing marginal productivity of capital). We also expect $C^o(q,K) < C^n(q,K)$ to reflect the higher costs of the new technology. As explained in Appendix A, the specific functional form that satisfies these conditions can be derived from the Cobb-Douglas production function:

$$C^j(q,K) = A_j \left[\frac{q}{K^{\alpha_j}}\right]^{1/\beta_j}, \tag{2.8}$$

where $\alpha_j \in (0,1)$ and $\beta_j \in (0,1)$ represent relative intensities of capital and other inputs and $A_j > 0$ is a generic scaling factor.

Further, we assume that the costs of new technology are affected by the learning:

$$\tilde{C}^n(q, K, x) = \left[\frac{x}{\bar{X}} + 1\right]^{\log_2 \gamma} C^n(q, K).$$

Table 2: Additional parameter values: Model III (capital)

Parameter	Unit	Notation	Value
Old tech. scale	\$B	A_o	36.52
New tech. scale	\$B	A_n	91.30
Old tech. capital exponent	-	α_o	0.2
New tech. capital exponent	-	α_n	0.2
Old tech. other exponent	-	β_o	0.8
New tech. other exponent	-	β_n	0.8
Old tech. initial capital	-	$\bar{K}_0^o$	1900
New tech. initial capital	-	$\bar{K}_0^n$	1
Depreciation rate	-	δ	0.1
Unconstrained emissions	tC	$Z_{\max}$	61.4698

Parametrization. The additional parameter values are listed in Table 2. The cost functions $C^j(q, K)$ and capital accumulation process $(\delta, f(I), K_0^j)$ are calibrated to match statistics of the basic model: $c^o(0) = 20$, $c^n(T) \approx 30$, $q^o(0) \approx 60$.

$K_0^n = 1$ is a token level to avoid division by zero in the cost function; setting K_0^o much higher implies that we start out with fully developed old technology. $Z_{\max}$ is changed because, despite our best efforts at calibrating the models to be similar, the different structure of the capital model implies a change to the optimal output level when emission are unconstrained (which is how z_T is determined).

Model. To keep the notation concise, we define several aggregate variables.

$$Q(t) = q^o(t) + q^n(t) \qquad C(t) = C^o(q^o(t), K^o(t)) + C^n(q^n(t), K^n(t))$$
$$I(t) = I^o(t) + I^n(t) \qquad K(t) = K^o(t) + K^n(t)$$

The new optimal control problem is given by

$$\underset{\{q^j,K^j,I^j,x,z\}(t)}{\text{maximize}} \quad \left\{ \int_0^T e^{-rt} \left[\tilde{S}(Q(t),t) - C(t) - K(t) \right] dt + e^{-rT} K(T) \right\} \tag{2.9a}$$

$$\text{subject to} \quad \dot{x}(t) = q^n(t), \qquad x(0) = x_0 = 0 \tag{2.9b}$$

$$\dot{K}^j(t) = -\delta K^j(t) + I^j(t), \qquad K^j(0) = \bar{K}_0^j \qquad j \in \{o,n\} \tag{2.9c}$$

$$\dot{z}(t) = e^{-at}[b_o q^o(t) + b_n q^n(t)], \qquad z(0) = z_0 = 0 \tag{2.9d}$$

$$z(T) \le z_T \tag{2.9e}$$

$$q^j(t) \ge 0, \qquad j \in \{o,n\} \tag{2.9f}$$

$$I^j(t) \ge 0, \qquad j \in \{o,n\}. \tag{2.9g}$$

Objective (2.9a) now computes the net welfare as the difference between consumer surplus and both production and investment costs. The new term $(e^{-rT}K(T))$ is used to avoid the terminal effects that arise because we use a finite-horizon approximation to an infinite-horizon problem. Without this correction, we would see the investment fall to zero as we approach $t = T$, causing a decline in capital and subsequent slowdown or decline in Q. The correction method we use can be interpreted as getting back the investment at the end of modeling period. This is not the only possible correction; see Barr and Manne (1967) and Lau et al. (2002) for a discussion of correction terms.

Constraint (2.9c) defines the law of motion for the capital—it is increased by investment and decreased by depreciation. Constraints (2.9d)–(2.9f) are identical to previous models; (2.9g) adds nonnegativity constraint on investment. While (2.8) requires $K^j(t) > 0$, we do not impose that constraint because (2.9c) and (2.9g) ensure that it holds.

3 Discretized Optimal Control Problem

In practice, we cannot expect to solve the optimal control problems like (2.5), (2.7), and (2.9) explicitly. We therefore approximate them by a finite-dimensional problem obtained by discretizing the time-dependent functions and collocating at a finite number of points. This approach results in NLPs that we can solve using large-scale optimization methods. The approach is discussed in detail by Betts (2001).

We discretize time $t \in [0,T]$ by $N+1$ equally spaced points $t_i = ih$, where $h := T/N$ is the step size and $i = 0,1,2,...,N$. Our convention is that q_i^n approximates $q^n(t)$ at t_i, namely, $q_i^n = q^n(t_i) = q^n(ih)$, with a similar convention for other variables.

Next, we discretize the differential equations such as (2.5b) by applying one of three Runge-Kutta methods and collocating at the discretization points,

$$x_{i+1} = x_i + hq_i^n, \qquad x_{i+1} = x_i + hq_{i+1}^n, \text{ and } \qquad x_{i+1} = x_i + h\frac{q_i^n + q_{i+1}^n}{2},$$

for the explicit Euler scheme, the implicit Euler scheme, and the trapezoidal method, respectively.

We note that the implicit Euler method and the trapezoidal method are implicit schemes. Since we are solving optimization problems, however, the complexity is not significantly increased. The Euler methods are both first-order accurate, while the trapezoidal rule is second-order accurate. For illustration, we use explicit Euler for Model I, implicit Euler for Model II, and trapezoidal for Model III. We note, however, that our AMPL models are set up so that each model can use any discretization.

3.1 Discretization of Model I (Basic)

The model variables are $\{q_i^o, q_i^n, x_i, z_i\}_{i=0}^N \in \mathbb{R}^{4(N+1)}$. To simplify the model statement, we define

$$\begin{aligned} W_i &= e^{-rhi}\left[\tilde{S}\{q_i^o + q_i^n, t\} - c_o q_i^o - c_n(x_i) q_i^n\right] \\ g_i &= e^{-ahi}[b_o q_i^o + b_n q_i^n]. \end{aligned}$$

With the explicit Euler discretization, the continuous problem (2.5) becomes

$$\begin{aligned} &\underset{q_i^o, q_i^n, x_i, z_i}{\text{maximize}} && h \sum_{i=0}^{N-1} W_i && \text{(3.1a)} \\ &\text{s.t.:} && x_{i+1} = x_i + h q_i^n, \quad i = 0, ..., N-1; \quad x_0 = 0 && \text{(3.1b)} \\ &&& z_{i+1} = z_i + h g_i, \quad i = 0, ..., N-1; \quad z_0 = 0 && \text{(3.1c)} \\ &&& z_N \le z_T && \text{(3.1d)} \\ &&& q_i^o \ge 0, \quad q_i^n \ge 0, \quad i = 0, ..., N. && \text{(3.1e)} \end{aligned}$$

3.2 Discretization of Model II (Adjustment Cost)

The model variables are $\{q_i^o, q_i^n, x_i, z_i, y_i\}_{i=0}^N \in \mathbb{R}^{5(N+1)}$, and the objective term becomes

$$W_i = e^{-ihr}\left\{S(q_i^o + q_i^n) - c_o q_i^o - c_n(x_i) q_i^n - c_a y_i^2\right\},$$

where y_i is the discretization of $y(t)$. The implicit Euler discretization of model (2.7) is given by

$$
\begin{aligned}
\underset{q_i^o, q_i^n, x_i, z_i, y_i}{\text{maximize}} \quad & h \sum_{i=0}^{N-1} W_i && (3.2a)\\
\text{s.t.:} \quad & x_{i+1} = x_i + h q_{i+1}^n, \qquad i = 0, ..., N-1; \qquad x_0 = 0 && (3.2b)\\
& q_{i+1}^n \le q_i^n + h y_{i+1} \qquad i = 0, ..., N-1; , \quad q_i^n = 0 && (3.2c)\\
& 0 \le y_i, \qquad i = 0, ..., N-1; && (3.2d)\\
& z_{i+1} = z_i + h g_{i+1}, \qquad i = 0, ..., N-1; \qquad z_0 = 0 && (3.2e)\\
& z_N \le z_T && (3.2f)\\
& q_i^o \ge 0, \qquad q_i^n \ge 0, \qquad i = 0, ..., N. && (3.2g)
\end{aligned}
$$

The differential inclusion (3.2c) is a new type of constraint, but it poses no difficulty for our optimization method.

3.3 Discretization of Model III (Capital Investment)

The model variables now are $\{q_i^j, I_i^j, K_i^j, x_i, z_i\}_{i=0}^N \in \mathbb{R}^{7(N+1)}$, and the new objective term is

$$
W_i = e^{-ihr} \left\{ S(q_i^o + q_i^n) - C^o(q_i^o, K_i^o) - C^n(q_i^n, K_i^n) - I_i^o - I_i^n \right\}.
$$

We also define

$$
k_i^j := -\delta K_i^j + I_i^j.
$$

The trapezoidal discretization of model (2.9) is

$$
\begin{aligned}
\underset{\{q_i^j, K_i^j, I_i^j, x_i, z_i\}}{\text{maximize}} \quad & \left\{ h \sum_{i=0}^{N-1} W_i + e^{-rNH}(K_N^o + K_N^n) \right\} && (3.3a)\\
\text{subject to} \quad & x_{i+1} = x_i + h \frac{q_i^n + q_{i+1}^n}{2} \qquad x_0 = 0 && (3.3b)\\
& K_{i+1}^j = K_i^j + h \frac{k_i^j + k_{i+1}^j}{2}, \qquad K_0^j = \bar{K}_0^j \qquad j \in \{o, n\} && (3.3c)\\
& z_{i+1} = z_i + h \frac{g_i + g_{i+1}}{2}, \qquad z_0 = 0 && (3.3d)\\
& z(T) \le z_T && (3.3e)\\
& q_i^j \ge 0, \qquad j \in \{o, n\} && (3.3f)\\
& I_i^j \ge 0, \qquad j \in \{o, n\}. && (3.3g)
\end{aligned}
$$

4 Numerical Solution and Results

All models are coded in the AMPL modeling language (Fourer et al., 2003). and solved by using KNITRO (Byrd et al., 2006). Unless specified otherwise, we use trapezoidal discretization with $h = 0.1$. The AMPL code is available online and is briefly described in Appendix B.

Several advantages accrue from attacking the problem directly as an optimal control problem, instead of following the Hamiltonian formalism. For example, we can easily include bounds on the controls in the optimal control problem. Further, tackling the optimization directly gives the numerical algorithms a better chance at finding maximizers instead of minimizers, or other stationary points, that are not distinguishable from the first-order conditions.

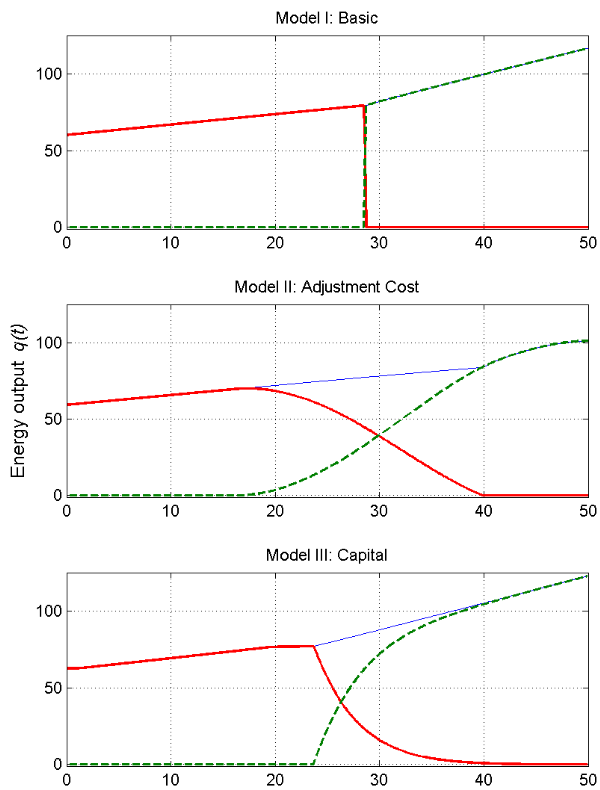

Figure 1: Optimal energy output schedules $(q^o(t), q^n(t))$ for the three models. Emission reduction rate $\zeta = 0.5$.

Figure 1 presents the optimal output quantities $(q^j(t))$ for our three models. We achieve transition in all cases, in the sense that $q^n(0) = 0$ and $q^o(T) = 0$. The common parametrization of demand ensures that $q^o(0)$ is similar for all models. Looking at $q^n(T)$, we see that the addition of adjustment costs in Model II predictably decreases it; Model III is not directly comparable to the other models.

Our key interest is the shape of the output schedules. Model I shows an instant transition; that is, there is no point in time where both $q^o(t)$ and $q^n(t)$ are positive. This is clearly not a realistic model. Model II has a continuous and smooth S-shaped transition, which is one of the common shapes of transition paths observed in the past (Jensen, 1982). Model III has a continuous transition with a kink at the start, followed by a concave segment, which is another historic shape described by Jensen (1982). Growth in total output is virtually unaffected by the transition, and appears to be linear, which is consistent with what we have seen in Model I.

Table 3: Computation time for various discretization methods and period lengths (h).

	Model I			Model II			Model III		
Step length h	1.0	0.5	0.1	1.0	0.5	0.1	1.0	0.5	0.1
Number of variables	200	400	2000	249	499	2499	400	800	4000
Number of constraints	100	200	1000	149	299	1499	202	402	2002
Computation time, s									
Explicit Euler	0.040	0.168	3.487	0.141	0.443	7.191	0.052	0.105	0.642
Implicit Euler	0.092	0.460	3.483	0.143	0.422	8.191	0.053	0.108	0.647
Trapezoidal	0.058	0.196	3.841	0.218	0.445	10.400	0.060	0.119	0.684

Table 3 shows the computation time for the three models and various discretization methods and step lengths. We see that step length is the main determining factor of the computation time. This result is expected because h determines the number of periods in the discretization scheme ($N = T/h$) and hence the number of variables and constraints. Given the same h, the discretization method does not matter nearly as much; moreover, no one method turns out to be the fastest across all step lengths. In the remainder of the section, we discuss the results of models I and III.

4.1 Discussion of Model I (Basic)

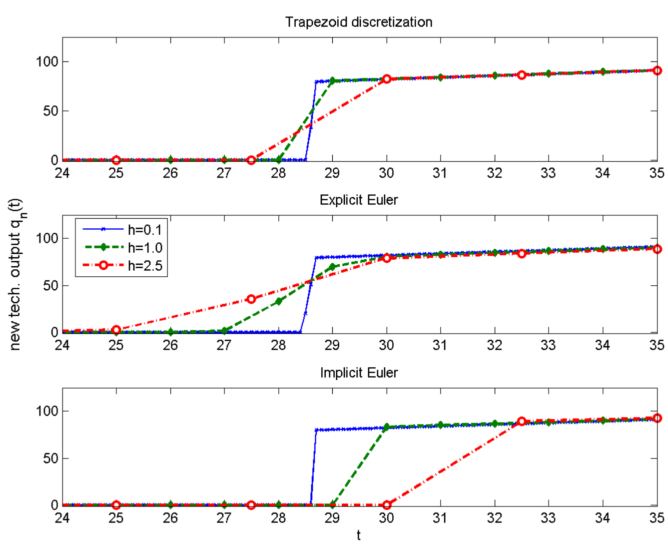

Figure 2: Model I (Basic). Optimal output for the new technology ($q^n(t)$) for various discretization methods and period lengths (h).

Figure 2 presents the solution to Model I given different discretizations methods and period lengths (h). We present the output only for the new technology, because the output for the old technology evolves in a symmetric fashion.

Clearly, the "smooth" transition that we observe for larger period lengths is an artifact of the discretization. As $h \to 0$, the solution approaches the instantaneous transition, which corresponds to the solution of the continuous-time model (2.5). We note that Euler methods carry a discretization error proportional to $O(h)$, making solutions with $h = 1$ or coarser time-steps questionable.

Further, the explicit Euler method results in qualitative error; while the trapezoidal and implicit Euler methods achieve transition in a single time-step, the explicit Explicit method takes 2–3 periods, creating the illusion of a smooth transition. The explicit Euler method stands out because the two other discretization methods are both implicit. Economically, an implicit method means that the firms begin to enjoy the benefits of learning before the period is over; that is q_i^n affects the total cost $(c_n(x_i)q_i^n)$ both directly and through x_i. Numerically, we prefer the use of implicit methods because they are better suited to stiff differential equations and differential algebraic equations that are found in optimal control problems.

4.2 Discussion of Model III (Capital)

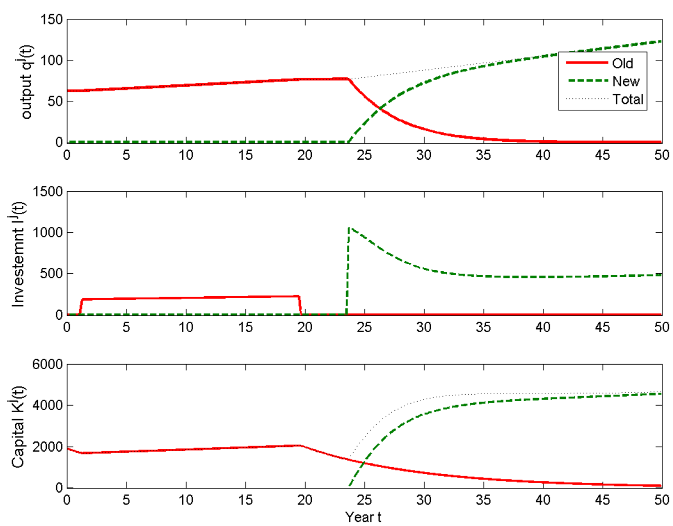

Figure 3: Model III (Capital). Optimal output, investment, and capital given 50% reduction in emissions.

Figure 3 presents the detailed solution to Model III, which is substantially different from the other ones. The total energy output continues to increase throughout the transition from old to new technology. The finite-horizon correction successfully eliminates terminal effects.

The investment schedule is discontinuous for both technologies, which seems reasonable, unlike the jump in outputs. Prior to the start of transition, there is period of zero investment into both

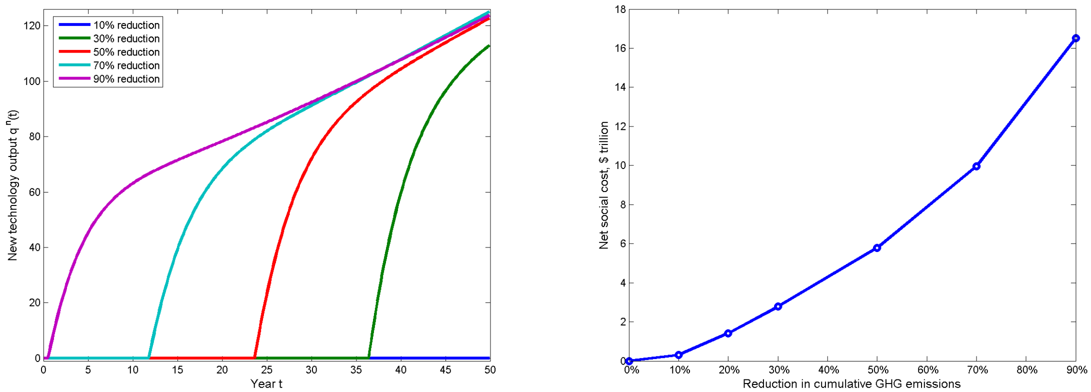

Figure 4: Model III (Capital). Left: the optimal $q^n(t)$ for various reductions targets ζ. Right: the corresponding abatement costs. the x-axis plots ζ, the y-axis the reduction in maximized objective from the case of $\zeta = 0$.

technologies, as the model waits for the old capital to depreciate before replacing it with the new technology. Once the transition starts, investment into the new technology spikes as the model tries to ramp up output to reap benefits of learning by doing.

The ease of computing the solutions allows us to study the impact of parameter values. For example, we can investigate the sensitivity to the emission reduction percentage by changing ζ illustrated in Figure 4. The left panel shows that smaller emission reductions require lower use of new technology and permit the transition to happen later. The right panel shows the social cost of reducing emissions. The convex shape is consistent with the increasing marginal costs commonly observed in economics.

5 Comments and Extensions

The primary goal of this paper is to develop solution methods for use in policy work. To do this, we will need to expand the model to include realistic features such as additional technologies and capital conversion. Our modeling approach allows us to add new technologies with minimal effort. For example, Figure 5 presents the solution to the capital model that includes a third technology, nuclear fusion, that is emission free but even more expensive ($A_3 > A_n$). We observe that industry first transitions to the intermediate technology but eventually adopts the cleanest one.

Another direction for further development of the model is the combination of learning by doing, transition costs, and capital accumulation in a single model but in a more industry-specific way. As Woerlen (2004) noted, learning-by-doing effects apply to the production of equipment, rather than to the energy generation itself. This fact suggests that experience should reduce investment rather than production cost, replacing I^j in (2.9a) with an investment cost function $f(I^j, K^j)$ that is increasing in I but decreasing in K.

Further, it is reasonable to allow the (costly) conversion of capital from one technology to the

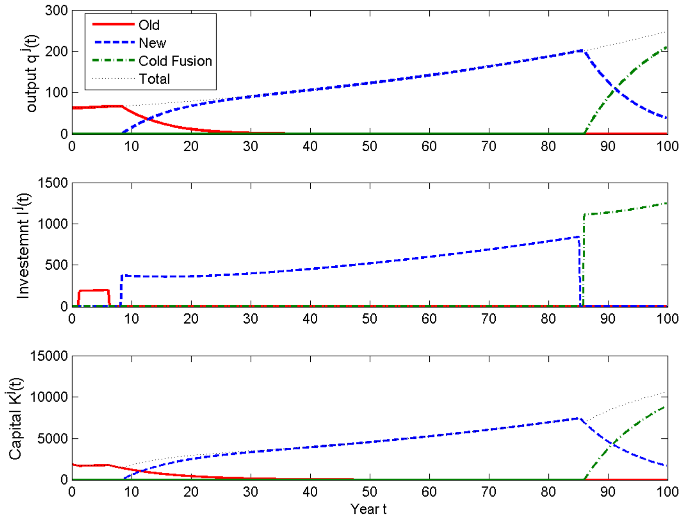

Figure 5: Capital model III with three technologies. Optimal output, investment, and capital given 80% reduction in emissions.

other (e.g., from fossil to biofuel). To avoid instant conversion, we could constrain the conversion rate from above or make the conversion costs convex. We might, however, need to take into account the shutdown time necessary to perform the conversion.

We note that our solutions have few areas of large curvature and are relatively smooth elsewhere. On such smooth segments, we can achieve adequate precision with relatively large periods, which means fewer variables and increased computation speed. But we still need a small step length during the transition period. Adaptive mesh refinement methods have been developed to automatically identify the areas where extra precision is needed. Betts (2001) describes how the computation time can be reduced by a factor of 2.

Acknowledgments

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract No. DE-AC02-06CH11357.

References

Balcer, Y. and Lippman, S. (1984). Technological expectations and adoption of improved technology. *Journal of Economic Theory*, 34:292–318.

Barr, J. R. and Manne, A. S. (1967). Numerical experiments with finite horizon planning models. *Indian Economic Review*, 2:1–29.

Benhabib, J. and Spiegel, M. (2005). Human capital and technology diffusion. *Handbook of Economic Growth*, 1A:935–966.

Betts, J. (2001). *Practical Methods for Optimal Control Using Nonlinear Programming*. Advances in Design and Control. SIAM, Philadelphia.

Byrd, R. H., Nocedal, J., and Waltz, R. A. (2006). KNITRO: An integrated package for nonlinear optimization. In di Pillo, G. and Roma, M., editors, *Large-Scale Nonlinear Optimization*, pages 35–59. Springer-Verlag, Berlin.

Cabral, L. (1990). On the adoption of innovations with network externalities. *Mathematical Social Sciences*, 19:299–308.

Fourer, R., Gay, D. M., and Kernighan, B. W. (2003). *AMPL: A Modelling Language for Mathematical Programming*. Books/Cole—Thomson Learning, 2nd edition.

Geroski, P. (2000). Models of technology diffusion. *Research Policy*, 29:603–625.

Jensen, R. (1982). Adoption and diffusion of an innovation of uncertain profitability. *Journal of Economic Theory*, 3:182–193.

Jensen, R. (1983). Innovation adoption and diffusion when there are competing innovations. *Journal of Economic Theory*, 29:167–171.

Lau, M. I., Pahlke, A., and Rutherford, T. F. (2002). Approximating infinite-horizon models in a complementarity format: A primer in dynamic general equilibrium analysis. *Journal of Economic Dynamics and Control*, 26(4):577–509.

Milliman, S. and Price, R. (1989). Firm incentives to promote technological change in pollution control. *Journal of Environmental Economics and Management*, 17:247–265.

Newell, R., Jaffe, A., and Stavins, R. (2006). The effects of economic and policy incentives on carbon mitigation technologies. *Energy Economics*, 28:563–578.

Popp, D. (2004). Innovation and energy prices. *Encyclopedia of Energy*, 3:451–458.

Tonks, I. (1986). The demand for information and the diffusion of a new product. *International Journal of Industrial Organization*, 4:397–408.

Woerlen, C. (2004). Experience curves for energy technologies. *Encyclopedia of Energy*, 2:641–649.

A Functional Forms and Model Parameters

Demand and Consumer Surplus $S(q)$ is derived from CES demand, which reduces to either a power function (with negative exponential) or a log function. We use an indefinite integral to compute surplus; neither formulation would be defined if we took a definite integral from 0 to Q.

Demand Growth is represented by (2.1) above; here we show its derivation. We assume that the demand growth takes the form

$$\tilde{q}(p,t) = e^{bt} q(p),$$

where $q(p)$ represents demand at $t = 0$, and is a monotone decreasing function. Hence, there exist inverse demand functions $p(q)$ and $\tilde{p}(q,t)$ that are related as follows:

$$\tilde{p}(q,t) = p(q/e^{bt}).$$

Consumer surplus is defined as an indefinite integral of inverse demand

$$\tilde{S}(q,t) = \int \tilde{p}(q,t)dq = \int p(q/e^{bt})dq = e^{bt} \int p(q/e^{bt})d(q/e^{bt}) = e^{bt} S(q/e^{bt}) + C,$$

where $S(q/e^{bt}) = \int p(q)dq$ and C is a constant of integration. Since $\tilde{S}$ is part of the objective, the constant does not affect the solution of the optimization problem.

Unit Cost Function $c_n(x)$ (also called "learning curve") is given by

$$c_n(x) = c_n^0 \left[\frac{x}{\bar{X}} + 1 \right]^{\log_2 \gamma}. \tag{A.1}$$

The interpretation of the exponent is that whenever experience doubles, $c(x)$ is multiplied by γ. The constant c_n^0 represents the cost at the top of the learning curve ($x = 0$). Since scale of demand would make optimal q_n's (and resulting x's) relatively large, the unit size of experience $\bar{X}$ scales x. The scaled experience is shifted by unity to ensure that $c_n(x)$ is decreasing for $x \geq 0$.

Total Cost Function (2.8) is used with the capital model (2.9). Let us assume that the energy production process uses two types of inputs—K units of capital and B units of other inputs (e.g., materials). The Cobb-Douglass production function translates these into output amount q:

$$q = DK^\alpha B^\beta. \tag{A.2}$$

Our model accounts for costs of capital via investment, so any additional cost of production is associated with other input B. Assume its price is p_B. Then the cost of producing q is

$$\begin{aligned} C(q,K) &= p_B B \\ &= A \left[\frac{q}{K^\alpha} \right]^{1/\beta}, \end{aligned}$$

where $A = p_B D^{-1/\beta}$.

B Description of the AMPL Code

The AMPL code for all three models is available at `www.mcs.anl.gov/~leyffer/OptTechPen`. In this appendix, we briefly describe the code for Model I and use it to highlight key features of AMPL language. The code consists of several files.

Z_run.ampl is the "run file" that incorporates all commands necessary to set up the model, solve it, and produce output. The run file also allows the user to change the discretization method and period length. To run it, one types `ampl Z_run.ampl` at the command prompt.

Z_optpen.mod is the main model file that sets the parameter values, declares model variables, and defines the constraints and other relationships that are independent of the discretization method.

Z_EE.mod, Z_EI.mod, Z_Tr.mod are three additional model files describing the model components specific to the explicit Euler, implicit Euler, and trapezoidal discretization, respectively. Only one of these files should be used; the user can change the discretization method by (un)commenting lines in `Z_run.ampl`.

Unlike programming languages (C, Fortran), which concentrate on computing outputs from inputs, modeling languages such as AMPL define the algebraic relationships between variables; the model is then submitted with a single command to one of many available solvers. AMPL has several types of statements; we illustrate them using examples of code from Model I.

Parameters are the constants that cannot be changed by the solver. One can set parameters equal to specific numbers, for example $T = 50$ and $N = 500$, or compute them from other parameters, such as $h = T/N$.

```
param tf := 50;
param hn := 500;
param  h := tf/nh;
```

Variables can be declared along with constraints on them or default values, such as the outputs $(q_i^j, i = 1, ..., N)$, which are declared to be nonnegative and given a default value of zero. Variables can also be declared to be a function of other variables and parameters, thus creating an implicit constraint, such as computation of total output Q_i.

```
var qo {0..nh} >=0, :=0;
var qn {0..nh} >=0, :=0;
var Q {i in 0..nh} = qo[i] + qn[i];
```

Objective The objective (3.1a) is represented as

```
maximize DiscWelf: sum {i in 0..nh-1}  h*welf[i];
```

where `welf[i]` stands in for W_i and `DiscWelf` is the variable that receives the maximized value of objective.

Constraints are directly specified. For example, the constraint (3.1b) becomes

```
subject to ode_x {i in 0..(nh-1)}:  x[i+1] = x[i] + h * qn[i];
```

where `ode_x` is the name assigned to that constraint (we could use it, for example, to retrieve the shadow cost of this constraint).